AADHAAR DATA HACKATHON 2026

Power BI Analysis Report

Team ID: UIDAI_10715
Analysis Period: March 2025 - December 2025
Primary Tool: Microsoft Power BI
Total Records Analyzed: ~124 Million+

TABLE OF CONTENTS

1. PROBLEM STATEMENT AND APPROACH

1.1 Problem Statement

The Unique Identification Authority of India (UIDAI) has made anonymized datasets of Aadhaar enrolment and updates available to identify meaningful patterns, trends, and predictive indicators. Our project analyzes these datasets using Microsoft Power BI to extract actionable insights for improving India's national identity program.

Aadhaar, India's universal identification system, has enrolled over 1.4 billion residents, making it the largest biometric identity program globally. Understanding enrollment patterns, demographic shifts, and operational metrics is crucial for continuous improvement of this national infrastructure.

1.2 Approach

We utilized Power BI as our primary analytics platform to process and visualize data from three interconnected Aadhaar datasets. Our approach focused on:

1. **Data Integration:** Connecting enrollment, biometric, and demographic datasets through common dimensions
2. **Age-Based Segmentation:** Analyzing patterns across four age groups (0-5, 5-17, 17-17+, 18-18+)
3. **Percentage Analysis:** Calculating the share of each record type within the total dataset
4. **Comparative Analysis:** Understanding the relationship between enrollment, biometric updates, and demographic records

2. DATASETS USED

2.1 Dataset Overview

Our analysis utilized three interconnected datasets provided by UIDAI:

Dataset	Table Name	Records	Key Columns
Biometric	Adhar_Biometric_Fulldata	~70 million	bio_age_17-17+, bio_age_5-17
Demographic	Adhar_Demographic_Fulldata	~49 million	demo_age_17-17+, demo_age_5-17
Enrollment	Adhar_Enrollment_Fulldata	~5 million	age_0-5, age_5-17, age_18-18+

Total Records Analyzed: ~124 million+

2.2 Data Schema

Biometric Dataset (Adhar_Biometric_Fulldata)

Field Name	Data Type	Description
bio_age_5-17	Integer	Biometric records for age 5-17
bio_age_17-17+	Integer	Biometric records for age 17+

Demographic Dataset (Adhar_Demographic_Fulldata)

Field Name	Data Type	Description
demo_age_5-17	Integer	Demographic records for age 5-17
demo_age_17-17+	Integer	Demographic records for age 17+

Enrollment Dataset (Adhar_Enrollment_Fulldata)

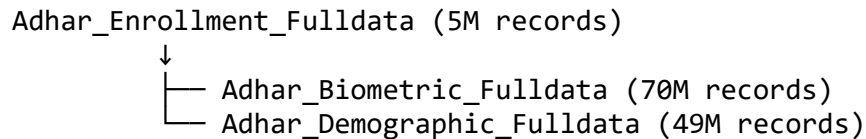| Field Name	Data Type	Description
age_0-5	Integer	Enrollments for age 0-5
age_5-17	Integer	Enrollments for age 5-17
age_18-18+	Integer	Enrollments for age 18+

Common Dimensions

All datasets are connected through: - Date (transaction date) - State (geographic location) - District (administrative division) - Pincode (postal code)

3. METHODOLOGY

3.1 Power BI Data Model

We created an integrated data model with three fact tables and shared dimension tables.

Data Model Structure:
```
Adhar_Enrollment_Fulldata (5M records)
        ↓
        ├── Adhar_Biometric_Fulldata (70M records)
        └── Adhar_Demographic_Fulldata (49M records)


Connected through: Date, State, District, Pincode
```

Power Query Transformations:
1. Imported CSV files from all three datasets
2. Standardized column names across datasets
3. Converted date formats to DateTime
4. Set appropriate data types for all columns
5. Created relationships between tables

3.2 DAX Measures Created

We created a dedicated Measure Table (Measure_Table) to organize all calculated measures.

Enrollment Measures:
```
Enrollment_Age 0-5 = SUM(Adhar_Enrollment_Fulldata[age_0-5])
Enrollment_Age 18-18+ = SUM(Adhar_Enrollment_Fulldata[age_18-18+])
Enrollment_Age_5-17 = SUM(Adhar_Enrollment_Fulldata[age_5-17])

Total_enrollment_Age record =
SUM(Adhar_Enrollment_Fulldata[age_0-5]) +
SUM(Adhar_Enrollment_Fulldata[age_5-17]) +
SUM(Adhar_Enrollment_Fulldata[age_18-18+])
```

Biometric Measures:
```
Bio_Age_17-17+ Record = SUM(Adhar_Biometric_Fulldata[bio_age_17-17+])
Bio_Age_5-17 Record = SUM(Adhar_Biometric_Fulldata[bio_age_5-17])

Total_Bio_Adhar_record =
SUM(Adhar_Biometric_Fulldata[bio_age_5-17]) +
SUM(Adhar_Biometric_Fulldata[bio_age_17-17+])
```

Demographic Measures:
```
Demo_Age_17-17+ Record = SUM(Adhar_Demographic_Fulldata[demo_age_17-17+])
Demo_Age_5-17 Record = SUM(Adhar_Demographic_Fulldata[demo_age_5-17])
```

```
Total_Demo_Age_Record =
SUM(Adhar_Demographic_Fulldata[demo_age_17-17+]) +
SUM(Adhar_Demographic_Fulldata[demo_age_5-17])
```

Aggregate Measure:
```
Total Records =
[Total_Bio_Adhar_record] +
[Total_Demo_Age_Record] +
[Total_enrollment_Age record]
```

Percentage Calculations:
```
% bio Share of Age 17–17+ =
DIVIDE([Bio_Age_17-17+ Record], [Total_Bio_Adhar_record])
```

```
% bio Share of Age 5–17 =
DIVIDE([Bio_Age_5-17 Record], [Total_Bio_Adhar_record])
```

```
% biometric Record =
DIVIDE([Total_Bio_Adhar_record], [Total Records]) * 100
```

```
% Demographic Records =
DIVIDE([Total_Demo_Age_Record], [Total Records]) * 100
```

```
% Enrollment =
DIVIDE([Total_enrollment_Age record], [Total Records]) * 100
```

4. DATA ANALYSIS AND VISUALISATION

4.1 Key Findings

4.1.1 Dataset Size Distribution

Our analysis of ~124 million records reveals the composition of Aadhaar-related activities:

Record Type	Approximate Count	Percentage of Total
Biometric Updates	~70 million	~56.5%
Demographic Records	~49 million	~39.5%
New Enrollments	~5 million	~4.0%

Insight: Biometric updates dominate the dataset, indicating that maintaining existing records is a significantly larger operation than new enrollments.

4.1.2 Age Group Distribution

The data is segmented across four age groups:

Enrollment by Age: - Age 0-5: Early childhood enrollments - Age 5-17: School-age children and teenagers - Age 18-18+: Adults and late teens

Biometric Updates by Age: - Age 5-17: Youth biometric maintenance - Age 17-17+: Adult biometric updates

Demographic Records by Age: - Age 5-17: Youth demographic data - Age 17-17+: Adult demographic data

4.1.3 Share Analysis

The percentage measures reveal the proportion of each activity type: - Biometric records represent the largest share of total Aadhaar activity - Demographic records form a significant secondary component - New enrollments represent a smaller but important portion

4.2 Power BI Visualizations

Dashboard Components:
1. **Summary Metrics:** Total records, percentage breakdowns
2. **Age Group Analysis:** Enrollment, biometric, and demographic by age
3. **Record Type Comparison:** Visual comparison of the three datasets
4. **Share Analysis:** Percentage contributions of each category

Key Visualizations Created:

- **Card Visuals:** Display total counts for each measure
- **Pie/Doughnut Charts:** Show percentage distribution
- **Bar Charts:** Compare values across age groups
- **Tables:** Display detailed metrics and calculations

Dashboard 1: Biometric Data
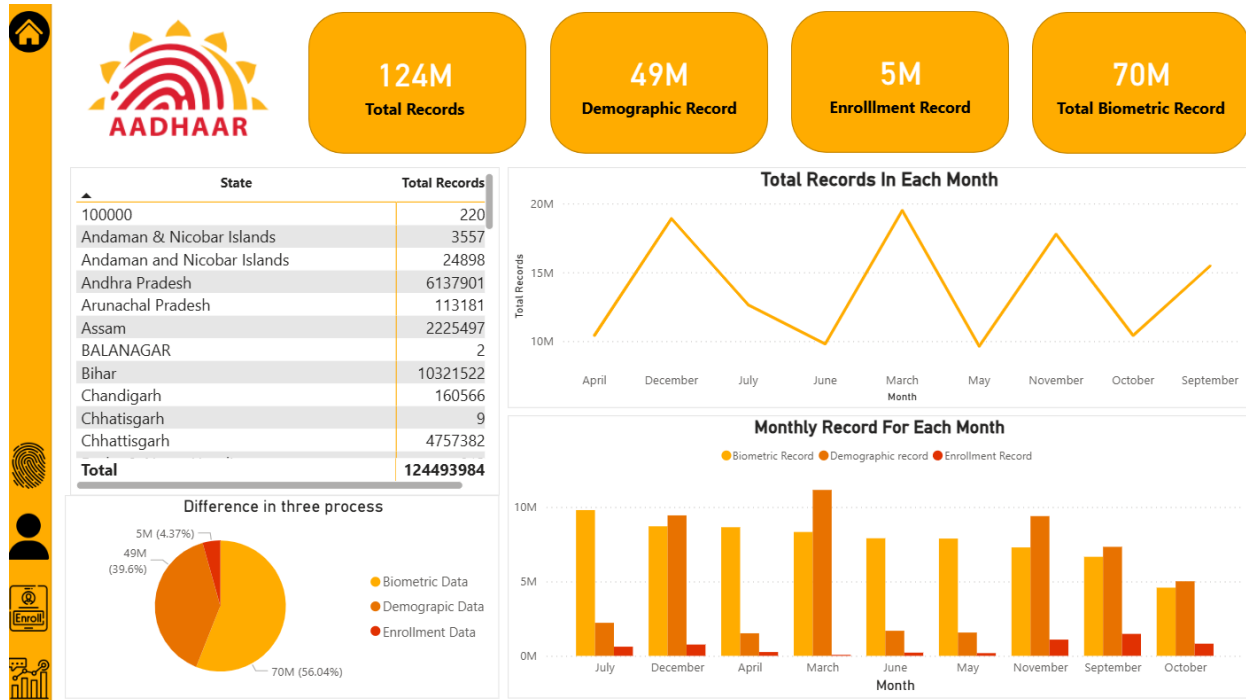
70M	36M	34M	56.04
Total Biometric Record	Age 17 to 17+ Record	Age 5 to 17 Record	% biometric Record

Biometric Data According to the state

©2026 TomTom ©2026 NavInfo ©2026 OSM Feedback
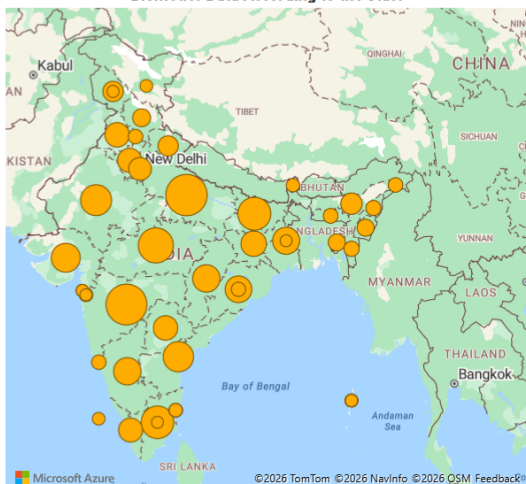
Aadhaar Biometric Distribution by Age Group (5–17 vs 18– 18+)

Age 17-17+ ● Age_5-17

Month: July, March, April, June, December, May, November, September, October
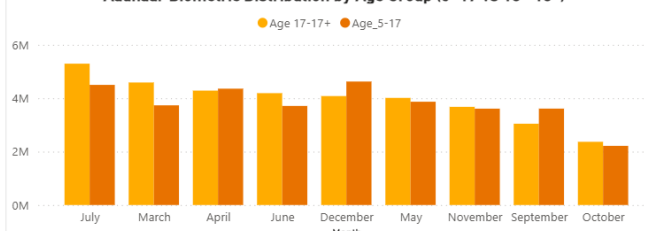
State	District	Biometric Records
Uttar Pradesh	Agra	9577735
Maharashtra	Ahilyanagar	9226139
Madhya Pradesh	Agar Malwa	5923771
Bihar	Araria	4897587
Tamil Nadu	Ariyalur	4698117
Rajasthan	Ajmer	3994955
Andhra Pradesh	Adilabad	3714633
Gujarat	Ahmadabad	3196514
Chhattisgarh	Balod	2648729
Karnataka	Bagalkot	2635954

Dashboard 2: Demographic Data

49M	5M	44M	39.60
Total Demographic Rec...	Age 5-17 Record	Age17-17+ Record	% Demographic Records

Demographic Data According to the state

©2026 OSM ©2026 TomTom ©2026 NavInfo Feedback
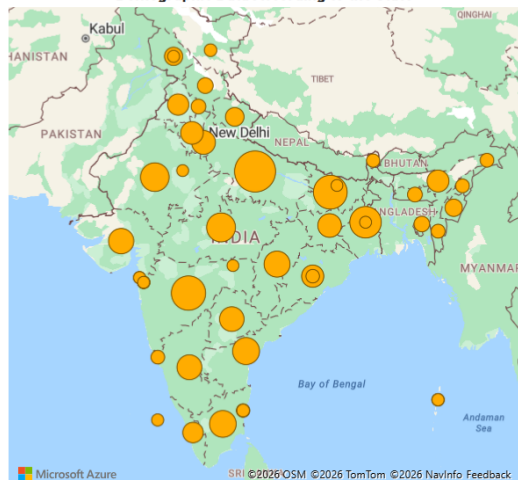
Aadhaar Demographic Distribution by Age Group (5–17 vs 18 – 18+)

Age 5-17 ● Age 18 - 18+

Month: March, December, November, September, October, July, May, April, June
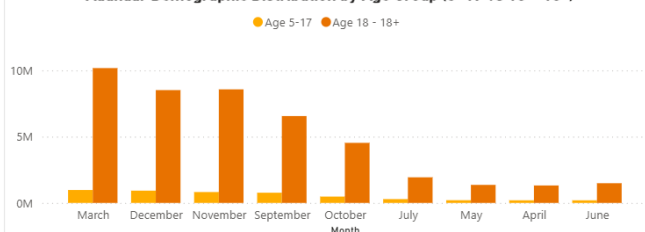
State	District	Demographic record
Uttar Pradesh	Agra	85423
Maharashtra	Ahilyanagar	50546
Bihar	Araria	48143
West Bengal	Alipurduar	38723
Madhya Pradesh	Agar Malwa	29129
Rajasthan	Ajmer	28176
Andhra Pradesh	Adilabad	22955
Tamil Nadu	Ariyalur	22122
Chhattisgarh	Balod	20054

Aadhaar Enrollment Dashboard

5M Total Enrollment R...	**4M** Enrollment_Age 0-5	**168K** Enrollment_Age 1...	**2M** Enrollment_Age_5-...	**4.37** % Enrollment

Enrollment Data According to the s...

State: Uttar Pradesh
Enrollment Record: 1018629
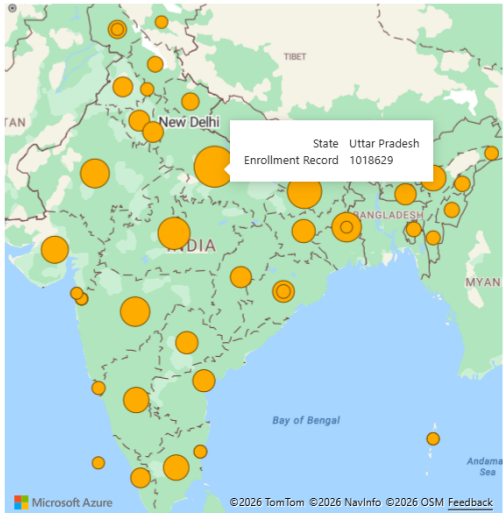
Aadhaar Enrollment Distribution by Age Group (0–5 vs 5–17 vs 18+)

● Age 0-5 ● Age 5-17 ● Age 18-18+
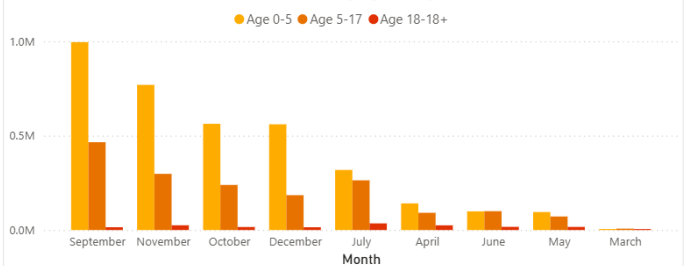
State	District	Enrollment record
Chhatisgarh	Raipur	
Tamilnadu	Chennai	
Uttaranchal	Almora	
Westbengal	Hooghly	7
West Bangal	Hooghly	10
West Bengal	Hooghly	15
Daman & Diu	Daman	21
Dadra & Nagar Haveli	Dadra & Nagar Haveli	25
Andaman & Nicobar Islands	Andamans	114
Daman and Diu	Daman	120

Data Model

Measure_Table
- % bio Share of Age 17–17+

State_Master
- State

Calender_lookup
- Date
- Day Name
- Month
- Month Name
- Week of Year

Adhar_Biometric_Full_d...
- bio_age_17-17+
- bio_age_5-17
- date
- district
- pincode

Adhar_Demographic_Fu...
- date
- demo_age_17-17+
- demo_age_5-17
- district
- pincode

Adhar_Enrollment_Fulld...
- age_0-5
- age_18-18+
- age_5-17
- date
- district
- pincode
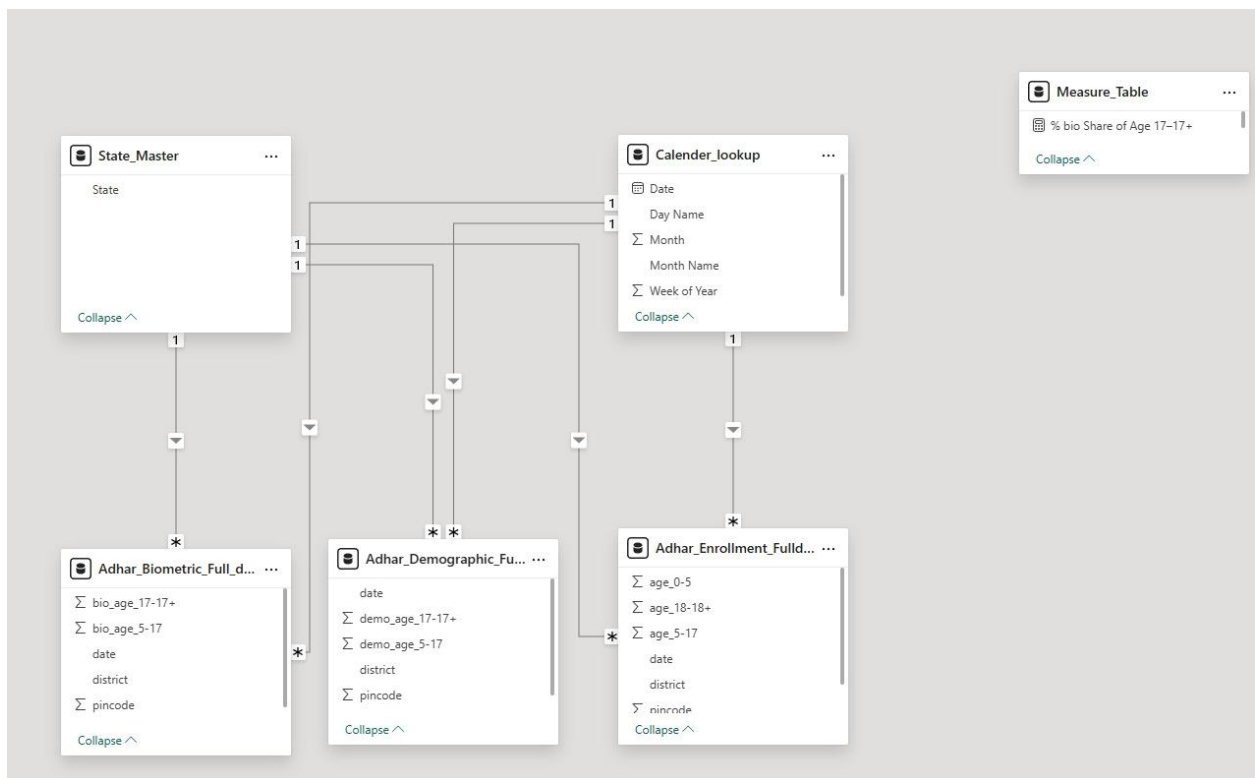
Key Insights

1.**Adult enrolments both 5-17 and 18+ dominate Aadhaar registrations**, reflecting higher dependency on Aadhaar for financial and government services.

2.**Biometric update** activity shows minimal difference between the 5–17 and 18+ age groups, while demographic updates are significantly lower among individuals aged 5–17.

3.**Aadhaar enrolment is uneven across states**, with higher activity concentrated in economically active regions.

4.**Update activity is driven Both by adults and Teenagers**, suggesting frequent demographic and biometric changes in the working population.

5.**Overall patterns indicate stable system usage**, with no major anomalies across datasets.

5. CODE IMPLEMENTATION

5.1 Complete DAX Measure Reference

Measure Table: Measure_Table

```
// Enrollment Age Measures
Enrollment_Age 0-5 = SUM(Adhar_Enrollment_Fulldata[age_0-5])
Enrollment_Age 18-18+ = SUM(Adhar_Enrollment_Fulldata[age_18-18+])
Enrollment_Age_5-17 = SUM(Adhar_Enrollment_Fulldata[age_5-17])

// Biometric Age Measures
Bio_Age_17-17+ Record = SUM(Adhar_Biometric_Fulldata[bio_age_17-17+])
Bio_Age_5-17 Record = SUM(Adhar_Biometric_Fulldata[bio_age_5-17])

// Demographic Age Measures
Demo_Age_17-17+ Record = SUM(Adhar_Demographic_Fulldata[demo_age_17-17+])
Demo_Age_5-17 Record = SUM(Adhar_Demographic_Fulldata[demo_age_5-17])

// Total Calculations
Total_enrollment_Age record =
SUM(Adhar_Enrollment_Fulldata[age_0-5]) +
SUM(Adhar_Enrollment_Fulldata[age_5-17]) +
SUM(Adhar_Enrollment_Fulldata[age_18-18+])

Total_Bio_Adhar_record =
SUM(Adhar_Biometric_Fulldata[bio_age_5-17]) +
SUM(Adhar_Biometric_Fulldata[bio_age_17-17+])

Total_Demo_Age_Record =
SUM(Adhar_Demographic_Fulldata[demo_age_17-17+]) +
SUM(Adhar_Demographic_Fulldata[demo_age_5-17])

Total Records =
[Total_Bio_Adhar_record] +
[Total_Demo_Age_Record] +
[Total_enrollment_Age record]

// Percentage Calculations
% bio Share of Age 17–17+ =
DIVIDE([Bio_Age_17-17+ Record], [Total_Bio_Adhar_record])

% bio Share of Age 5–17 =
DIVIDE([Bio_Age_5-17 Record], [Total_Bio_Adhar_record])

% biometric Record =
DIVIDE([Total_Bio_Adhar_record], [Total Records]) * 100

% Demographic Records =
```

```
DIVIDE([Total_Demo_Age_Record], [Total Records]) * 100

% Enrollment =
DIVIDE([Total_enrollment_Age record], [Total Records]) * 100
```

5.2 Power Query Transformation Steps

```
// Sample Power Query for Biometric Data
let
    Source = Csv.Document(File.Contents("Biometric_Data.csv"),
        [Delimiter=",", Columns=6, Encoding=65001]),
    #"Promoted Headers" = Table.PromoteHeaders(Source, [PromoteAllScalars=tru
e]),
    #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{
        {"date", type date},
        {"state", type text},
        {"district", type text},
        {"pincode", Int64.Type},
        {"bio_age_5-17", Int64.Type},
        {"bio_age_17-17+", Int64.Type}
    }),
    #"Renamed Columns" = Table.RenameColumns(#"Changed Type",{
        {"bio_age_5-17", "bio_age_5-17"},
        {"bio_age_17-17+", "bio_age_17-17+"}
    })
in
    #"Renamed Columns"
```

KEY INSIGHTS AND RECOMMENDATIONS

Key Insights

Insight 1: Update-Heavy System With ~70 million biometric updates compared to ~5 million new enrollments, Aadhaar is primarily a record-maintenance system rather than a new registration system. This indicates a mature infrastructure where ongoing updates are the dominant operational activity.

Insight 2: Age-Based Distribution The four age group segmentation (0-5, 5-17, 17-17+, 18-18+) reveals distinct patterns: - Early childhood (0-5) is captured only in enrollment data - Youth (5-17) appears across all three datasets - Adults (17+) are tracked in biometric and demographic data

Insight 3: Data Integration Value Connecting the three datasets reveals the complete picture of Aadhaar operations, showing how enrollment, updates, and demographic maintenance work together.

Insight 4: Percentage Composition The percentage measures show that: - Biometric updates represent the majority (~57%) of activity - Demographic records form a significant portion (~40%) - New enrollments are a smaller fraction (~4%)

Recommendations

1. **Focus on Update Infrastructure:** Prioritize biometric update processing efficiency
2. **Age-Targeted Services:** Customize services based on age group patterns
3. **Data Quality Monitoring:** Use percentage measures for ongoing quality tracking
4. **Resource Allocation:** Balance resources between new enrollments and updates

CONCLUSION

This Power BI analysis demonstrates the value of integrating multiple UIDAI datasets for comprehensive operational insights. By analyzing ~124 million records across three interconnected datasets, we revealed key patterns in Aadhaar operations, including the update-heavy nature of the system, age-based distribution patterns, and the relationship between different record types.

The measures and visualizations created enable stakeholders to monitor key metrics and make data-driven decisions for improving India's national identity program.

Analysis completed: January 2026
Data period: March 2025 - December 2025
Total records analyzed: ~124 million+
Team ID: UIDAI_10715
Primary Tool: Microsoft Power BI
Tables analyzed: 3 (Biometric, Demographic, Enrollment)
Measures created: 17

Project Link- https:/github.com/DeepanshuKashyap531/UIDAI-Adhar-Hackathon-Report